AD-A208 490 MENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. RE UNCLASSIFIED	1b. RESTRICTIVE MARKINGS NONE
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFIT/CI/CIA-88-215

6a. NAME OF PERFORMING ORGANIZATION AFIT STUDENT AT St Mary's University	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFIT/CIA
6c. ADDRESS (City, State, and ZIP Code)		7b. ADDRESS (City, State, and ZIP Code) Wright-Patterson AFB OH 45433-6583

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification) (UNCLASSIFIED)
Meta-Analysis of Armed Service Vocational Aptitude Battery Composite Validity Data

12. PERSONAL AUTHOR(S)
Nicole S. Stermer

13a. TYPE OF REPORT THESIS/DISSERTATION	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, Month, Day) 1988	15. PAGE COUNT 105

16. SUPPLEMENTARY NOTATION
APPROVED FOR PUBLIC RELEASE IAW AFR 190-1
ERNEST A. HAYGOOD, 1st Lt, USAF
Executive Officer, Civilian Institution Programs

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

DTIC
ELECTE
JUN 02 1989
S H D

89 6 02 012

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☑ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL ERNEST A. HAYGOOD, 1st Lt, USAF	22b. TELEPHONE (Include Area Code) (513) 255-2259	22c. OFFICE SYMBOL AFIT/CI

DD Form 1473, JUN 86   Previous editions are obsolete.   SECURITY CLASSIFICATION OF THIS PAGE

AFIT/CI "OVERPRINT"

META-ANALYSIS OF

ARMED SERVICE VOCATIONAL APTITUDE BATTERY

COMPOSITE VALIDITY DATA

A
THESIS

Presented to the Faculty of the Graduate School
of St. Mary's University in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE

in

Industrial Psychology

BY

Nicole S. Stermer, BA

San Antonio, Texas

December, 1988

ACKNOWLEDGEMENTS

I wish to thank my committee chairman, Dr. Malcolm Ree, for his ideas, excellent criticism, and consistent encouragement. He kept my train of thought on the right track.

I also wish to thank my committee members, Dr. Willie Silva and Dr. Jeff Kantor, for their help and support. They helped to make sure my train went in the right direction.

I particularly want to thank Greg, my husband, who provided the fuel that kept my train going.

META-ANALYSIS OF

ARMED SERVICE VOCATIONAL APTITUDE BATTERY

COMPOSITE VALIDITY DATA

Nicole S. Stermer

St. Mary's University, 1988

Supervising Professor: Dr. Malcolm James Ree

This study investigated the efficacy of selection procedures

associated with the Armed Service Vocational Aptitude Battery

(ASVAB). The main hypothesis tested was whether the Armed

Forces Qualification Test (AFQT), an ASVAB composite, is a valid

predictor of training success. Additional hypotheses investigated

whether the AFQT is a more valid predictor of training success

than the individual career-specific selector composites. A final

hypothesis dealt with the expectation that the AFQT would show a

larger improvement over selector composites in validities for

females than for males.

Results indicate the AFQT has high predictive validity for

all military occupations included in the sample. Differential

validity is observed to favor three of the four ASVAB composites

(Mechanical, General, and Electronic). For the fourth composite,

(Administrative), the AFQT has higher validity, and the increment

Cont

does favor females. Two possible explanations for the findings
are discussed. The need for continued research is indicated. (SDW)

(Thesis.)

TABLE OF CONTENTS

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

CHAPTER 1

INTRODUCTION

As a personnel selection instrument, the Armed Service
Vocational Aptitude Battery (ASVAB) is a vital component in the
maintenance of a quality military force. By investigating the
method in which ASVAB scores can be applied in selection
procedures, this study attempts to aid the continual research
efforts to maintain the ASVAB as a state-of-the-art selection
instrument.

To elucidate the scope of this research problem, a number
of issues involved in personnel selection are discussed,
including the concept of g, and gender differences in ability
measures. The ASVAB is then described in terms of its history,
content, reliability, and validity. Lastly, validity
generalization and meta-analysis are discussed, as the basis
for this study.

History of Personnel Selection

The issues surrounding the topic of personnel selection
are much more complex than would initially be assumed by a casual
observer. A historical review is required before delving into
this many-faceted topic.

Measurement of human abilities has long been of interest to

scientists. Sir Francis Galton set up an anthropometric laboratory in the late 18th century. He measured sensory acuity and reaction times of thousands of subjects, believing that these types of abilities gauged the level of an individual's intellect (Anastasi, 1982). More refined measures of intelligence began principally with Binet, who in 1905 developed an intelligence test based on what he felt were the essential components of intelligence. He believed intellect consists of more abstract faculties, such as judgement and comprehension. Binet's intelligence test was developed for the purpose of identifying the educable mentally retarded children in Paris schools (DuBois, 1970).

The first group test used to assess abilities particularly for the purpose of personnel selection was developed during World War I, by a team of psychologists headed by Robert Yerkes (Graham & Lilly, 1984). The resulting test called the Army Alpha was the first multiple choice, objectively scorable, group administerable test devised. As such, it provided the military with a highly efficient means of measuring cognitive ability, or intelligence, of World War I recruits. Though the Army Alpha was developed to reflect a measure of general intelligence, it did contain individual subtests (specifically, Oral Directions, Arithmetrical Reasoning, Practical Judgement, Synonym-Antonym, Disarranged Sentences, Number Series Completion, Analogies, and

Information).

At that time in the history of psychological testing, "the prevailing opinion was that man possesses a single, all-important intellectual ability" (Ghiselli, 1973). Charles Spearman (1904, 1927) was a leading pioneer in the investigation of the concept of intellectual ability. Through his invention of the procedure of factor analysis, Spearman found evidence for his Two-Factor Theory of intelligence. His factor analytic procedures identified intercorrelations among tests of intellectual ability. He defined the commonality among the tests as representing general intelligence, labeled g, which composed the first factor of his theory. The second factor of his theory, s, represented the many factors specific to each test or type of test.

Faith in the existence of Spearman's g faded following Thurstone's (1938) application of his centroid factor method. Thurstone did not locate g, or any common factor, in a battery of 56 tests. Upon this evidence, Thurstone developed his theory of Primary Mental Abilities, which claimed that seven unique factors exist, each representing a separate mental ability, and that no single index could satisfactorily explain the concept of intellectual ability.

Thurstone's rejection of the existence of g did not stand for long. Spearman reviewed Thurstone's procedures, and by reanalyzing the data, did locate the general factor. In addition,

when Thurstone attempted to devise tests to measure his primary factors, he found that the primaries were intercorrelated, rather than independent as he had hypothesized (McNemar, 1964).

Though Thurstone devoted many years to the task of searching for pure measures of distinct abilities, he was unable to produce tests which remained uncorrelated with one another when based on a large, representative population (Jensen, 1986). This phenomenon, known as Positive Manifold, shows that mental ability tests with adequate reliabilities, when administered to representative populations, will always result in positive intercorrelations. Thurstone finally admitted that a general factor was required to explain the intercorrelations among his primary factors (Thurstone & Thurstone, 1941).

Despite the preliminary evidence in support of Spearman's g, the general movement in test development began to lean toward measurement of specific abilities, downplaying the concept of g.

Between World Wars I and II, psychologists hypothesized that batteries of tests measuring specific aptitudes could produce a composite score that would comprise an aptitude test tailored to a particular job or curriculum. Hunter (1984) stated this hypothesis was not tested until recently, and fails in all but a few special cases. Though untested, the concept of differential aptitude testing became particularly appealing to the military, due to rapid advances in military technology. During World War

II, many leading psychometricians applied multiple regression
strategies to aptitude batteries in order to optimally predict
performance in technical military career fields. Multiple or
differential aptitude batteries were employed in the civilian
sector as well.

Validation results of such batteries have led to an
emphasis on quantitative and verbal abilities, and also on
technical aptitudes particularly for military batteries. The
three most commonly used differential aptitude batteries at
present are: the ASVAB, developed by the Department of Defense
(DoD), (U.S. DoD, 1984); the General Aptitude Test Battery
(GATB), developed by the U.S. Employment Service (USES, 1970);
and the Differential Aptitude Test (DAT), published by The
Psychological Corporation.

The ASVAB

The ASVAB was developed in the 1960's for the purpose of
selection and classification of United States military personnel.
Previously, the Army, Navy, Marines, and Air Force used test
batteries that were developed separately by each service.

The Selective Service Act of 1948 led to the development of
the Armed Forces Qualification Test (AFQT), the goal of which was
to promote a more equitable distribution of abilities among the
various services. The AFQT was developed in a joint-services

project, for administration to all military applicants.

Standardization of the AFQT against the Army's entrance exam,

the Army General Classification Test (AGCT) utilized test scores

of all military enlisted personnel as of December 31, 1944. This

group became a reference population for subsequent AFQT versions.

The AFQT has since been used to compare the distribution of

abilities of accessions among the services, to screen applicants

for enlistment, and to report accession quality to the U.S.

Congress.

In 1958, the Airman Qualifying Examination (AQE) was

introduced into high schools by the Air Force for the purpose of

assisting in occupational counseling. Then in 1966, DoD

initiated development of a single test to be used for selection

and classification by all services. This joint-services project

resulted in the ASVAB. Form 1 of the ASVAB replaced the AQE for

high school use in 1968, and it was replaced by alternate form 2

in 1973. Alternate form 3 was used for Air Force recruiting

beginning in 1973, and by the Marines in 1975.

Early successes of the ASVAB led to its use as the sole

instrument for selection, classification, and assignment of all

service personnel. In 1976, ASVAB form 5 was provided to high

schools, while parallel forms 6 and 7 were provided to Military

Entrance Processing Stations (MEPSs), thus replacing the

batteries used by the individual services.

The Air Force Human Resources Laboratory (AFHRL) is currently the lead laboratory responsible for continuing ASVAB research, development, and validation. ASVAB forms 8, 9, and 10, which were developed to provide more accurate measures at lower levels of ability, were placed in use in October, 1980. Replacement forms 11, 12, 13, and 14, developed to be parallel to forms 8, 9, and 10, began use in October, 1984, and are the forms currently used.

The 1980 Profile of American Youth, a DoD-sponsored study, provided a more current reference population against which ASVAB scores could be interpreted. A nationally representative sample of about 12,000 men and women, aged 16 to 23, took form 8AX (identical to form 8A) of the ASVAB. Bock and Mislevy (1981) provide supporting data as to the representativeness of the sample, and the accuracy of the test and procedures used.

The content of the ASVAB forms 8 through 14 consists of 10 subtests: General Science (GS), Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), Numerical Operations (NO), Coding Speed (CS), Auto and Shop Information (AS), Mathematics Knowledge (MK), Mechanical Comprehension (MC), and Electronics Information (EI). The subject areas have been chosen according to their abilities to predict success in military training courses.

The current AFQT used by all services consists of a

composite of four ASVAB subtests (AR, WK, PC, and a half-
weighting of NO; see Table 1). The AFQT is used for initial
selection or rejection of applicants. Each of the services
employs additional subtest configurations for classification
purposes. The Air Force utilizes four such composites:
Mechanical (M), Administrative (A), General (G), and Electronic
(E). (See Table 1 for compositions of these composites.) The
composites, or aptitude indexes (AIs), were developed through
multiple regression procedures (Hunter, Crosson, & Friedman,
1985).

In the recruitment process, an applicant is administered a
current form of the ASVAB. His/her AFQT score, if sufficiently
high, qualifies the individual for enlistment. The AI scores
are then used, in addition to manning requirements and the
enlistee's preferences, to place the individual in a training
program for a specific job, or in a career field following
successful completion of Basic Military Training. Some highly
selective Air Force career fields employ higher AI cutoff scores
than others.

Studies of ASVAB reliabilities have yielded results
indicating that the ASVAB subtests have satisfactory reliability.
Ree, Mullins, Mathews, and Massey (1982) computed internal
consistency reliabilities for the eight power subtests for ASVAB
8a, 8b, 9a, 9b, 10a, and 10b. The average subtest reliabilities

Table 1

ASVAB Subtest Composites: Forms 8 Through 14

ASVAB Subtest	N of Items	AFQT	M[a]	A[b]	G[c]	E[d]
General Science (GS)	25		X			X
Arithmetic Reasoning (AR)	30	X			X	X
Word Knowledge (WK)	35	X		X	X	
Paragraph Comprehension (PC)	15	X		X	X	
Numerical Operations (NO)[e] [f]	50	X		X		
Coding Speed (CS)[e]	84			X		
Auto and Shop Information (AS)[g]	25		X			
Mathematics Knowledge (MK)	25					X
Mechanical Comprehension (MC)	25		X			
Electronics Information (EI)	20					X

[a] M = Mechanical Composite

[b] A = Administrative Composite

[c] G = General Composite

[d] E = Electronic Composite

[e] NO and CS are speeded subtests. All others are power subtests.

[f] NO is weighted at $\frac{1}{2}$ for AFQT inclusion.

[g] AS is double-weighted for inclusion in the M composite.

across all forms ranged from .81 to .92. Reliabilities of the two speeded subtests (NO and CS) were indirectly inferred from subtest intercorrelations, which were observed to range from .53 to .70 (U.S. DoD, 1984).

Hunter, Crosson, and Friedman (1985) reported reliabilities for all 10 subtests as computed by extrapolation from previous data (Friedman, Streicher, Wing, Grafton, & Mitchell, 1983; Kass, Mitchell, Grafton, & Wing, 1982). Hunter and his associates utilized a reliability theory formula which forecasts estimates from one population to another which has a different variance, thereby attempting to show what the ASVAB form 8, 9, and 10 subtest correlations would have been, were there no error of measurement. Corrected reliabilities thus computed ranged from .66 to .85.

Although internal consistency reliabilities provide useful information, parallel-forms (or alternate-forms) reliabilities are preferred. Parallel-forms methods correspond to the classic definition of reliability as the ratio of true-score variance to observed-score variance. Such methods avoid the possible bias caused by time effects, as encountered in test-retest methods. Parallel-forms methods also may be used with quickly-paced tests (such as the two speeded subtests of the ASVAB, NO and CS), whereas internal consistency methods will overestimate the reliabilities of such tests.

Palmer, Hartke, Ree, Welsh, and Valentine (1988) computed parallel-forms reliabilities for subtests and composites of ASVAB forms 8, 9, 10, and 11. (The results apply to forms 8 through 14, as all these forms are parallel). Their results, based on data from a sample of 75,000 armed service applicants, indicate that for each subtest, reliabilities are similar across all forms. The coefficients observed for the subtests, across all forms, ranged from .67 to .88. The composite coefficients (including the AIs and the AFQT) ranged from .87 to .93. Reliability coefficients were observed to be higher in general for males and whites than for females, blacks, or hispanics, although the reliabilities for females, blacks, and hispanics remained adequate. Reliability coefficients for females ranged from .56 to .92; for blacks, from .83 to .90; for hispanics, from .80 to .90.

This compilation of studies indicates stable, satisfactory levels of reliability across ASVAB forms, subtests, and composites.

The statistical validity of the ASVAB is an estimation of how useful it is as a tool to predict job performance in military career fields. Job performance measures have not, however, proven to be available criteria for estimating ASVAB validities, as there are no measures which can be commonly applied across all military occupations. However, technical training course

grades have been used for several decades.

ASVAB validity coefficients are computed by correlating the test scores with training school performance measures. Training school grades provide objective measures, and are obtainable across all occupations. Furthermore, the content of training programs is established according to job performance requirements. Within the Air Force, occupational analyses based on the Instructional Systems Development (ISD) process are performed for each career field. These analyses utilize the Comprehensive Occupational Data Analysis Program (CODAP) to ensure that training school content is based upon required job knowledge (Stacey, Weismuller, Barton, & Rogers, 1974). Hunter (1984) has stated that job knowledge and job performance measures are very highly correlated:

> Cognitive ability proved to be a very good predictor of job knowledge and is, thus, a good predictor of job performance. To a lesser extent, cognitive ability also predicts job performance directly . . . . data using work-sample tests show that there is a very high correlation between job knowledge and job performance. People who do not know much about the job will perform poorly. The multiple regression of work sample performance onto ability and knowledge shows that it is job knowledge which has the larger direct impact on performance. (p 52 and 53)

Thus not only do job knowledge measures accurately predict job performance, but cognitive ability measures, such as technical school grades, are highly appropriate for assessing job knowledge.

Each of the military services conducted validation studies for ASVAB forms 8, 9, and 10, in 1983. Detailed results and conclusions from these studies are presented in the ASVAB Test Manual (U.S. DoD, 1984). For all services, reported validities are sufficiently strong to predict training success. For Air Force data, no major validity differences were seen between black and white or male and female subgroups, for the specialties which had adequate-sized samples and for which validities reached useful levels.

Numerous extraneous variables can effect the measurement of the validity of occupational tests. Determination of ASVAB validities is complicated by the effects of range restriction in both the lower and upper score ranges. Applicants in the low end of the scale are eliminated from consideration, and thus from the validation sample. Likewise, those whose scores fall above certain cutoff scores may be selected for certain exclusive career fields, thereby also contributing to a restriction in range. Corrections for range restriction may be applied, and will be discussed in conjunction with meta-analytic procedures.

Additional variables which may effect ASVAB validity data are the differences among training course difficulty, course

content, and grading systems. Courses can vary from weeks to months in length, and from basic to highly technical skills. Impacting on these variables is the number of students per course. Some longer, more technical courses require a lengthy time period to develop an adequate-sized sample for validation purposes. Also, variations in grading systems exist, such as pass/fail systems versus letter grades, which may impact on validity measures.

Personnel Selection

Employee selection procedures, including recruiting, interviewing, testing, and validating selection criteria cost organizations billions of dollars per year. Choice of selection methods, and validation of those methods are thus of vital concern to employers. In addition to ensuring selection of quality personnel, employers must ensure their selection procedures are within the requirements of federal guidelines (Uniform Guidelines on Employee Selection Procedures, 1978). The guidelines are provided with the intention of promoting equal employment opportunities among all persons, but specifically for certain protected groups. The guidelines require validation studies for each occupation for which a selection instrument is used, if that instrument adversely effects hiring of the specified protected groups.

While attempting to avoid adverse impact in hiring, the guidelines urge employers to seek selection procedures alternative to tests, but that are equally as valid. That standardized tests are valid methods of personnel selection has been sufficiently supported by Ghiselli (1966, 1973), who reviewed hundreds of criterion-related validity studies. He concluded that "for every job, there is at least one type of test which has at least moderate validity" (Ghiselli, 1973, p 477-478). Attempts to validate alternative selection methods, however, have not met with such success. Reilly and Chao (1982) reviewed research on the validities of eight categories of selection methods: 1. biographical data (biodata), 2. interviews, 3. peer evaluation, 4. self-assessment, 5. reference checks, 6. academic performance measures, 7. expert judgement, and 8. projective techniques. They concluded:

> Only biodata and peer evaluation were supported as having validities substantially equal to those for standardized tests . . . . data, where available, offered no clear indication that any of the alternatives met the criterion of having equal validity with less adverse impact. (p 1)

Hunter and Hunter (1984) assessed validities of various predictors, with training performance determined through supervisor ratings (N ranged from 1,789 to 32,124). Ability composites provided the highest average validity, of .53,

followed by job tryouts at .44, and biodata at .37. The

remaining eight predictors ranged in validity from -.01 to .26.

The Uniform Guidelines' test validation requirements are

based on the belief that a test which has demonstrated validity

in one situation or group will not necessarily have acceptable

validity in others. This exemplifies the theory of situational

specificity, which is the contention that although jobs may

appear very similar, there are subtle, yet important differences

in jobs, tasks, individuals, or identifiable subgroups of

individuals which moderate the predictor-criterion relationship.

For example, a test found to adequately predict performance of

Chicago policemen might, due to racial or geographic differences,

or other unknown and unspecifiable differences, be invalid for

San Francisco policemen.

Prevalence of the theory of situational specificity resulted

in numerous validation studies for small groups, and for nearly

identical settings. The validity generalization work of Schmidt

and Hunter and their associates has shown this theory to be

questionable (Schmidt & Hunter, 1977; Schmidt, Hunter,

Pearlman, & Shane, 1979; Schmidt, Hunter, & Urry, 1976). They

contend that observed differences in validities across similar

situations and jobs are almost entirely due to statistical

artifacts. Their results have consistently indicated that test

validities are highly transportable across groups and situations.

Societal consequences of the importance of intelligence (or g) and the ability to measure it have definite implications regarding personnel selection procedures. The research has shown that a g factor exists, that it is relevant to job performance, and that it can be assessed by use of standardized tests. What, then, are the societal implications of utilizing tests as selection techniques? Two major theories, Functionalist and Revisionist, have argued that differences in intelligence are of little or no importance in the work place.

The Functionalist theorists assume that education provides job-relevant knowledge required for diverse occupations. They believe also that differences in job performance result primarily from specific skills that have been learned, not from a general ability, and that access to education is not equally obtainable for all individuals. Functionalist reforms would focus on revising educational policies by providing more equal access for underprivelaged groups (Gottfredson, 1986).

Revisionist beliefs are similar, yet are more extreme. They argue that an occupational hierarchy is unnecessary, and promotes unjustifiable differences in rewards. According to Gottfredson (1986):

> The revisionist perspective argues, moreover, that the high differentiation and specialization of work activities in our society, as well as the accompanying large differences

in the entrance requirements and socioeconomic rewards
among occupations, are the result of capitalists
intentionally fragmenting and "de-skilling" work in order
to increase not the efficiency of work, but their control
over workers. (p 382)

The solution to the problem, as the Revisionists see it, is to
restructure the jobs themselves, or the personnel selection
procedures, rather than revise education or training. They
assume the outcome to be c   .ncreased equality and productivity.

These theorists fail to consider the most basic concern of
the hiring organizations, namely, the need for obtaining
qualified personnel. It is extremely difficult for employers to
be able to predict each employee's exact worth. During the
hiring procedure, employers can only assume, at best, a level of
productivity, or worth, for specific individuals. Through
trial-and-error procedures, organizations have come to identify
indicators of applicants' potentials. Educational credentials
have come to be an extensively used indicator of success mainly
due to their moderately high correlation with intelligence
measures ($r = .60$) (Gottfredson, 1986).

Attempts at social revisions with the dual goals of
increasing both equality and productivity have met with little
success. A trade-off between the two is inevitable. Attempts to
increase equality by randomly assigning workers to differentially

g-loaded jobs would result in decreased productivity. Conversely, increasing the validity of job-person matching, though perceived fair ai ' equitable, would further exaggerate the present occupational hierarchy (Gottfredson, 1986).

A popular criticism of selection testing is that it promulgates racial and ethnic inequalities, by identifying as more qualified the individuals who are white, or are from middle class or higher socioeconomic backgrounds. Major supporters of this belief have been Ralph Nader and Allen Nairn (Nairn, 1980). Nader's criticism of the testing industry focused primarily on the Scholastic Aptitude Test (SAT), a college-entrance exam produced by the Educational Testing Service (ETS). Racial and ethnic criticisms against the use of standardized g-loaded tests have simply not held up technically. Research on this question has consistently revealed that ethnic groups' score differences on ability tests represent true, consistent differences between the groups. Spearman (1927) reported marked differences in mean scores of blacks and whites on a battery of 10 highly g-loaded tests. Jensen (1985) sustained and corroborated Spearman's findings, and reported that no studies had been fc ad to contradict them. The diverging distributions of g amcng races and genders will likely result in a proportionately lower number of blacks, hispanics, and women selected than whites and men, via cognitive ability measures.

That test results may produce ethnic disparity in numbers hired does not lessen the strength of the evidence for g, nor can it weaken the validity of the tests. The theory of differential vallidity, which contends that validities are different for various groups of people, has been repeatedly disconfirmed (Bartlett, Bobko, Mosier, & Hannon, 1978; Hunter, Schmidt, & Hunter, 1979; Schmidt, Pearlman, & Hunter, 1980). The model of test fairness adopted in the Uniform Guidelines on Employee Selection Procedures (1978) is the regression model, which defines a test as unfair if it predicts lower levels of job performance for a minority group than that group actually achieves. As stated by Schmidt and Hunter (1981) "The accumulated evidence on this theory is clear: Lower test scores among minorities are accompanied by lower job performance, exactly as in the case of the majority" (p 1131).

We have seen evidence for the diverging distributions of test scores of general cognitive ability measures for blacks versus whites. Can we expect similar results for men versus women?

Literature reviews (Jensen, 1980; Maccoby, 1966; Maccoby & Jacklin, 1974) have identified three areas in which gender differences appear: verbal, quantitative, and spatial abilities.

Girls appear to verbally outperform boys beginning at very early ages, perhaps even with their first words (McCarthy, 1954).

The differences are small, and boys begin to catch up until about age 10 or 11. After puberty however, "girls average close to a quarter of a standard deviation higher than boys on verbal tasks . . . . The sex difference in verbal ability after puberty appears to be a genuine phenomenon and not just a measurement artifact" (Jensen, 1980, p 625). Jensen's review of studies published since 1966 showed that, of 58 studies investigating general intelligence, 15 significantly favored females, whereas only 3 favored males. This is to be expected, in light of the knowledge that many general intelligence tests are heavily weighted on verbal ability, and require reading.

The pattern of quantitative ability is exactly the opposite of verbal ability. Boys are clearly superior as adolescents, and this difference remains throughout life. Differences of one-fifth to two-thirds of a standard deviation have been observed by the end of high school. This is again a true difference, and not an artifact of test bias. After equating males and females on the number of courses taken, males still emerge superior, suggesting that the difference is not due to greater training for males (Maccoby & Jacklin, 1974).

The third area of observed differences, spatial ability, has been difficult to define and investigate. Smith's (1964) definition was "the perception, retention, and recognition (or reproduction) of a figure or pattern in its correct proportions"

(p 96). Numerous skills have been proposed as relative to
spatial ability, including: auditory localization, size-
distance constancy, tactual recognition of objects as they
change in space, and matching pairs of items that are mirror-
reversals or that have been rotated in space. Despite the
difficulties encountered in clarifying this concept, researchers
have consistently indicated that males are superior on spatial
abilities, particularly after adolescence. Maccoby's (1966)
review showed an advantage for boys on the DAT and the Primary
Mental Abilities Test. Jensen (1980) stated that only about
one-fourth of the females exceed the male mean on spatial-
visualization tests. The reason for this difference is still
unclear. Proposed causal factors include maturation rates,
hormonal factors, sex-linkage, and cultural effects.

In general, females perform better than males on verbal
tasks, and somewhat better than males on general intelligence
tests, while males are superior on quantitative and spatial
tasks.

As in the case of racial differences, the gender differences
discovered do not adversely effect interpretations of test
validities. Differential predictive validity of tests by sex
has concentrated on prediction of college grades. In Jensen's
(1980) review, he located no studies reporting lower validities
for women. For the ASVAB, as previously discussed, no

significant differences were observed between validity data of males and females.

Another consideration regarding personnel selection is the cost. As stated previously, billions of dollars are spent for hiring programs each year. According to Hunter and Schmidt (1982), most major corporations abandoned the use of selection tests during the previous 10 years, in order to comply more fully with federal guidelines. During this same period, the rate of growth in productivity declined from three and a half percent per year, to zero and even below. The authors attributed this decline, at least in part, to decreased test usage. After dramatically lowering their testing standards to require only minimum scores (at approximately a seventh grade level), U.S. Steel's detailed training records showed that:

1. Scores on mastery tests given during training declined markedly;

2. the flunk-out and drop-out rates increased dramatically;

3. average training time and training costs for those who did make it through the program increased substantially, and

4. average ratings of later performance on the job declined. (Hunter & Schmidt, 1982, p 298).

Schmidt, Hunter, Outerbridge, & Trattner (1986) have empirically estimated the economic impact of various selection

methods for the federal work force. They determined that, for the federal government, a one-year cohort based on valid, cognitive ability measures, would increase productivity values up to six hundred million dollars for each year the employees are retained. Hunter and Schmidt (1983) estimated a labor cost savings of 15 to 20 percent for any organization, due solely to selection based on cognitive ability.

Without doubt, the legal, societal, and financial aspects of personnel selection must be of vital concern to any organization. According to John Hawk (1986), employees who are selected on the basis of validity generalization (based on the concept of g) are more productive, learn faster, are more quality conscious, and are more satisfied with their jobs. This results in lower turnover rates. Ability tests remain the most valid personnel selection procedure known. The future of personnel selection procedures appears to lie in the direction of improving testing programs, rather than in the search for alternatives to testing.

Validity Generalization and Meta-analysis

Schmidt and Hunter and their associates have extensively investigated the feasibility of transporting test validities across groups and situations. Such validity generalization research counters the claim of the situational specificity hypothesis. By generalizing, or "borrowing" validities,

researchers have shown that numerous, costly validation studies are unnecessary.

The concept of generalizing results across studies or situations is not new. For many years scientists have been combining results from numerous independent studies in order to infer a general conclusion. Integrative reviews have employed various methods, including vote counting (counting of positive and negative results), counting significant results, and averaging statistics across studies. Rosenthal (1978) describes a number of statistical combination procedures.

As a vanguard of the validity generalization movement, Glass (1976, 1977) devised a statistical method he entitled meta-analysis. His method combines the results of individual studies by converging the results into a common metric, coding various characteristics of the studies, then using conventional statistics to determine whether there is an overall effect.

Schmidt and Hunter and their associates have built a meta-analytic procedure based on the ideas of Glass, but incorporated some modifications. The Schmidt-Hunter model, utilizing principles of Bayesian statistics, improves on Glass's procedure by providing corrections for sources of error. Schmidt, Hunter, Pearlman, and Shane (1979) identified seven artifactual sources of variance:

1. Differences between studies in criterion reliability;

2. differences between studies in test reliability;

3. differences between studies in range restriction;

4. sampling error (variance due to $N < \infty$);

5. differences between studies in amount and kind of criterion contamination and deficiency;

6. computational and typographical error, and

7. slight differences in factor structure between tests of a given type. (p 260-261)

Schmidt, et al., do not correct for the last three sources of error, having determined that it is difficult to estimate their frequency or magnitude. Not correcting for these error sources results in conservative variance estimates. They also apply a conservative decision rule, which accounts for the fact that only four of the seven sources of artifactual variance are corrected for. According to their rule, they reject the situational specificity hypothesis if 75 percent or more of the variance of the validity coefficients is accounted for by the four corrected artifacts. If a large proportion of the observed variability is attributable to artifacts, the conclusion is that the true population variance is negligible, and validity coefficients can be generalized across situations.

A central concept in the Schmidt-Hunter methodology is that validation attempts have historically relied on insufficient sample sizes, and have thus led to erroneous conclusions,

because researchers have interpreted sampling error as true variance (Schmidt & Hunter, 1978). Many supposed moderators of predictor-criterion relationships, including race, ethnicity, sex, age, socioeconomic status, leadership style, and geographic area, have been identified through belief in the "law of small numbers." This law proposes that a small random sample can be considered as representative of a population as a large random sample. The error of this proposition can be shown using an example of race as a moderator. For validation studies, minority samples have been generally smaller than for the majority. In single-group studies, this produced a large number of white-significant, black-nonsignificant findings. This is another example of interpreting error variance (due to lack of statistical power) rather than true variance.

Schmidt and Hunter (1978) stated that the differential validity approach is sound, but requires an extremely large sample size in each group (in the hundreds) to have a .90 probability of detecting true differences. This is a recognition that statistical power increases with the square of the sample size, and that large samples are necessary to produce powerful statistical tests. The small sample size problem encountered in single-group validity studies and differential validity approaches are common among job-test validations, because research must often be based on the number of available workers.

The meta-analytic approach combines the results of numerous small-sample studies, and thus provides overall results based on a sufficient sample.

Applications of Schmidt and Hunter's meta-analytic procedure has shown, in numerous studies, that much of the observed variation in validities for similar job-test combinations is artifactual. For example, Schmidt, Hunter, and Caplan (1981) investigated the transportability of validities of four types of cognitive tests. They reported that "support for generalizability was substantiated for general mental ability and arithmetic tests" (p 261). They found that sampling error alone accounted for 90 percent of all variance due to artifacts. Additional support is given by Pearlman, Schmidt, and Hunter (1980), Schmidt and Hunter (1984), and Schmidt, Hunter, Pearlman, and Shane (1979).

The evidence shows the situational specificity hypothesis to be invalid. For a given job-test combination, the four sources of error corrected for by meta-analytic procedures are capable of producing as much variance as is generally observed between validation studies.

Validity Generalization and the ASVAB

It appears that a general ability, or g factor, serves as a link among validities of cognitive ability tests. Though authors disagree on the proper definitions of g and intelligence, they

agree that differences among individuals in a general ability

factor are largely responsible for differences in success and

status in the United States (Tyler, 1986).

Though general cognitive ability is seen as a valuable

predictor of performance, with some validity for all jobs, there

are jobs for which its validity is relatively low. Data

available from U.S. Employment Service (USES) validation studies

on 515 jobs representative of those in the Dictionary of

Occupational Titles (DOT) point toward job complexity as a

moderator of validities (Hunter, Crosson, & Friedman, 1985).

Even in the worst cases, however, Hunter concluded an average

cognitive ability validity of .37 for training success and .32

for job proficiency. Additionally, Hunter et al. (1985)

analyzed existing Army data (Helme, Gibson, & Brogden, 1957),

and found indications of higher validity for cognitive ability

on complex tasks (such as decision making), accompanied by higher

validities for psychomotor ability measures on tasks of low

complexity (such as following instructions, as in "feeding" and

"off-bearing" tasks. Feeding refers to putting something into a

machine; off-bearing is removing something from a machine.)

Thus as job complexity varies, higher predictive validities may

be obtained by shifting between cognitive and psychomotor

abilities as the primary predictors.

If cognitive ability is such a pervasive concept, what then,

is its relationship to the ASVAB, which was developed to reflect
differential aptitudes? The theory of differential aptitudes
would propose that ASVAB subtest composite (AI) validities will
be high for jobs in corresponding occupational areas, and lower
for unrelated occupations. For specific jobs, poor
predictability of one aptitude could thus be offset by
emphasizing another aptitude with higher validity. Hunter,
et al. (1985) disbelieved this theory, and subsequently showed
it to be untrue. They analyzed ASVAB scores from samples of
military personnel (N ranged from 107 to 13,904) according to the
subtest composites used by high schools for occupational
counseling. Composite content was as follows: Verbal =
WK + PC + GS; Quantitative = AR + MK; Technical = AS + MC + EI;
Speed = NO + CS. Each occupational composite was found to be
nearly as valid for other occupational areas as it was for its
own occupations. A clear lack of differential validity was
observed.

In an additional analysis, Hunter et al. (1985) analyzed
an Air Force sample of 29,619 recruits, with AI validities
computed for 70 jobs (Wilbourn, Valentine, & Ree, 1984). After
the data had been corrected for restriction in range, Hunter
and his associates concluded that only in the electronics area
was the AI more valid for its corresponding jobs than for any of
the other composites. In the same study, mechanical and general

occupational performances were predicted better by composites other than the ones designed specifically for these career fields. Predictive ability of a General Cognitive Ability composite was found to be as high as those for Administrative and Electronic composites, and higher than those for the other two composites (Mechanical and General).

Viewing such results as these, we can see a strong indication of a General Cognitive Ability factor moderating the relationships among ASVAB validities. According to Jensen (1986):

> Virtually no one today disputes that a g factor can be extracted from the correlations among any large and diverse collection of mental ability tests, and that the g factor is usually substantiated in the sense of subsuming a relatively large proportion of the total variance in all of the tests as compared with other factors besides g. The point that is being questioned is whether the g factor represents any reality outside the operations of psychometric tests and factor analysis. (p 302)

Extensive evidence does support the existence of the g phenomenon beyond psychometric operations. Non-psychometric correlates of g have been identified in recent years. g loadings have been found related to heritability (Block, 1968; Tambs, Sundet, & Magnus, 1984), and to test correlations among family

members (Nagoshi & Johnson, 1986). A review by Jensen (1983) identified 12 studies providing evidence that g loadings are related to inbreeding depression. Other reviews (Eysenk & Barrett, 1985; Haier, Robinson, Braden, & Williams, 1983) have led to the conclusion that g correlates with evoked electrical brain potentials.

Jensen (1986) went on to describe a hierarchical factor structure wherein variance unique to individual tests is filtered out at the level of primary factors, variance unique to primary factors is filtered out at the level of secondary factors, etcetera, with the g factor appearing at the highest level. He states:

> The g factor is remarkably stable across different collections of mental tests, even collections of tests that bear hardly any superficial resemblance to one another . . . . all so-called intelligence tests, or "IQ" tests, even when they have not been constructed with reference to factor analysis, are found to be very highly g loaded. (p 308 and 310)

Though the ASVAB was not constructed via factor analytic methods, researchers have applied such procedures to the test to determine its factor structure. These analyses have indicated that the regression-based composites do not correspond to factor-analytic composites (Hunter, 1983b; U.S. Naval Personnel

Research Activity, 1981).

In Hunter's investigation (1983b) he produced a hierarchical factor model of military subtests through analysis of available factor-analyzed data (Maier & Grafton, 1981; Sims & Hiatt, 1981; Thorndike, 1957). He identified three factors common to all the data: Quantitative, Verbal, and Technical composites. The factor model identified by Hunter (see figure 1) exemplifies the hierarchical structure described by Jensen (1986). Because Hunter's extensive study clearly describes the relationships among ASVAB subtests, it will be discussed in some detail.

Hunter (1983b) investigated additional data (Kass, Mitchell, Grafton, & Wing, 1982), based on ASVAB forms 8, 9, and 10, using confirmatory factor analysis (N = 98,689). Only 6 of the 10 subtests were used. One subtest, AS, did not load significantly on any factor. The inclusion of AS in the Technical factor with EI and MC would have generated inconsistent data. For this reason, AS was omitted from further analysis. Hunter found NO and CS (the two speeded subtests) to correlate highly with each other, yet NO consistently correlated higher with other subtests than did CS. He found that, in the presence of good verbal and quantitative measures, NO did not raise validity meaningfully. Also, NO and CS did not appear to measure the same construct as Perceptual Speed subtests of other batteries, thus these subtests were excluded for the reasons specified. The last exclusion was

Figure 1

Hunter's Hierarchical Factor Model

General

Cognitive

Ability

Quantitative

Aptitude

AR   MK

Verbal

Aptitude

WK   GS

Technical

Aptitude

P.I   MC
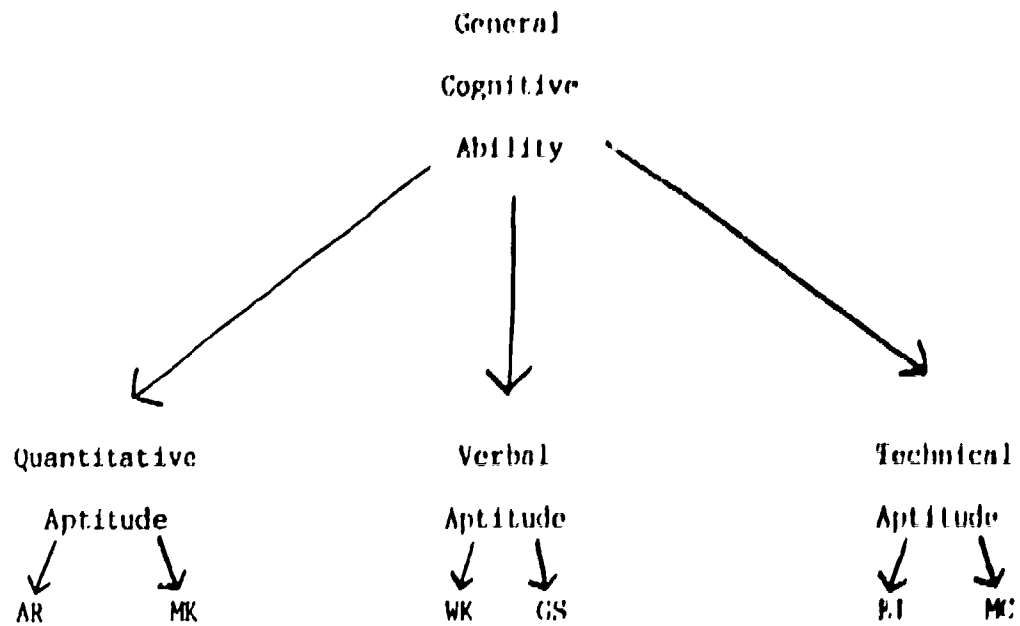
PC. Though very similar in composition to WK, PC had a
reliability of about .75, while the reliability of WK was found
to be .90. Weighting schemes for combining the two by giving
more weight to WK did not improve reliability, so PC was dropped.
Hunter stressed that these four subtests were not excluded for
reasons of invalidity, but for the purpose of allowing
construction of composites that are comparable with civilian
test battery data.

Hunter found that the remaining six subtests produced three
highly intercorrelated factors, designated as Quantitative
Aptitude (AR + MK), Verbal Aptitude (WK + GS), and Technical
Aptitude (EI + MC). His analysis showed that validity is not
higher when analyses are performed with all ten subtests. Table
2 presents the ASVAB composite structure, and the factor
structure isolated by Hunter. He concluded that "factor-analytic
composites are more valid than the multiple-regression composites
for predicting job performance" (Hunter, 1984, p 72).

Through a subsequent confirmatory factor analysis, Hunter
identified a second-order factor which he labeled General
Cognitive Ability. The best estimate of this general factor was
defined as the sum of the three aptitude composites (Verbal +
Quantitative + Technical), or as the sum of the six underlying
subtests. Table 2 shows the composition of General Cognitive
Ability.

Table 2

ASVAB Composite and Factor Structures

	AS[a]	PC[a]	Speed[b]		General Cognitive Ability					
			NO[a]	CS[a]	Quant[b c]		Verbal[b c]		Tech[b c]	
					AR	MK	WK	GS	EI	MC
M	X							X		X
A		X	X	X			X			
G		X			X		X			
E					X	X		X	X	
AFQT		X	X		X		X			

[a] Indicates subtests excluded from Hunter's model.

[b] Indicates the four factors initially derived through exploratory factor analysis.

[c] Indicates the three factors isolated by confirmatory factor analysis.

In yet another confirmatory factor analysis from the same study, Hunter analyzed the data available from Sims and Hiatt (1981) and Maier and Grafton (1981), both derived from ASVAB forms 8, 9, and 10, and he also included data from Thorndike (1957), based on the now-obsolete Army AC-1B. The Maier and Grafton, and Sims and Hiatt data produced a similar pattern (see figure 2) but the Thorndike data produced two differences relative to the model. An additional factor, Perceptual Aptitude, was obtained consisting of now-obsolete subtests. Also, two of the AC-1B primary factors contain subtest configurations different from the other data (due to test revisions over the years), thus these subtest validities are not shown in figure 2.

That all the data sources produced a similar pattern of aptitudes provides strong support for existence of the second-order factor, General Cognitive Ability. For example, results from the Kass, et al. (1982) data indicate that the correlations between the aptitude factors differ by no more than sampling error (see Table 3). Hunter's measurement model provides evidence that the individual ASVAB subtests contribute to overall validity via their contributions to one of the three primary aptitudes. The validities of the primary aptitudes, in turn, are not useful due to independent contributions, but to their inclusion in the second-order factor, General Cognitive Ability.

Figure 2

Path Model Relative to Military Composite Data

General Cognitive Ability

.90[a] .90[a] .90[a] .62[a]
.93[b] .93[b] .93[b] .63[b]
.86[c] .86[c] .86[c] .66[c]

Quantitative Aptitude   Verbal Aptitude   Technical Aptitude   Job Performance

.82[a] .82[a]   .81[a] .81[a]   .78[a] .78[a]
.87[b] .87[b]   .86[b] .86[b]   .84[b] .84[b]

MK   AR   WK   GS   EI   MC

---
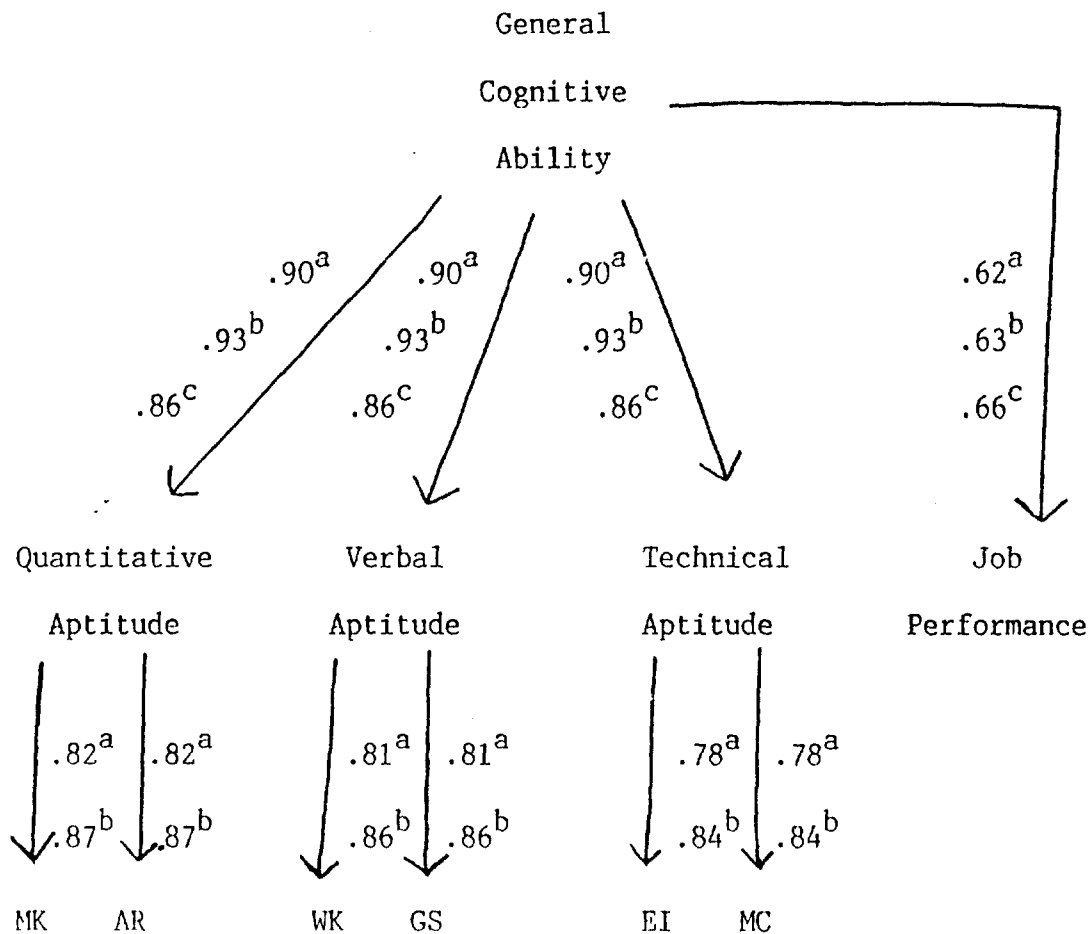
[a]   Sims & Hiatt (1981) data.

[b]   Maier & Grafton (1981) data.

[c]   Thorndike (1957) data.

Table 3

Intercorrelations of Factors

Aptitude Factors	Q^a	V^b	T^c	G^d
Q	1.00			
V	.82	1.00		
T	.80	.89	1.00	
G	.86	.96	.93	1.00

Note. Data as analyzed by Hunter, et al. (1985), obtained from Kass, et al. (1982). (N = 98,689)

[a] Q (Quantitative Aptitude) = AR + MK

[b] V (Verbal Aptitude) = WK + GS

[c] T (Technical Aptitude) = EI + MC

[d] G (General Cognitive Ability)

The general factor provides a measure of job performance due to its relationship to job knowledge acquisition.

General Cognitive Ability was seen to predict job training performance in specific areas as well as or better than the regression-based composite of specific abilities designed expressly for that occupational area. Only one exception was observed. Perceptual Speed made an incremental contribution to validity over General Cognitive Ability in predicting performance for clerical occupations (Hunter, 1984). Hunter presumed this was due to effects of the speeded ASVAB subtest, CS. His meta-analysis of eight military studies (1985) verified the evidence for differential validity in business and clerical occupations due to the Perceptual Speed factor. The validity of Speed for clerical work was .43, versus .37 for other types of work. Hunter identified a general Perceptual Aptitude factor as improving validities beyond the influence of General Cognitive Ability relative to Thorndike's (1957) data. Current versions of the ASVAB however, do not contain Perceptual Aptitude measures.

Hunter's extensive work on military data has shown the ASVAB composites to be of acceptable validity, though in light of General Cognitive Ability as a moderating factor, perhaps the composites are of questionable utility.

Though Hunter and Schmidt's validity generalization procedure has been shown to be a viable procedure, it has not

been without critics. Forty questions regarding meta-analysis and its applications to validity generalization are presented in a debate format (Sackett, Schmitt, Tenopyr, Kehoe, & Zedeck, 1985; Schmidt, Hunter, Pearlman, & Hirsch, 1985). For 23 of the questions, Schmidt, et al.'s responses stand as acceptable. Schmidt (1988) comments:

> The commentators . . . . took no issue with the major practical conclusion of meta-analytic research in personnel selection: that validities, particularly of cognitive tests, have been shown to be widely generalizable across settings, jobs, populations, organizations, geographical areas, time periods, etc. Most of the commentary was in the nature of attempts to "fine tune" statistical methods or applications. (p 179)

Practical acceptance of meta-analysis is evident from the wide application it has been given, including topics outside the realm of personnel selection. Examples include: correlates of role conflict and role ambiguity (Fisher & Gittelson, 1983), evaluation of Fiedler's theory of leadership (Premack & Wanous, 1985), and ability of financial analysts to predict stock growth (Coggin & Hunter, 1983).

The USES and the federal government are currently the two largest users of validity generalization. The USES has adopted validity generalization as the basis for its testing program

that operates through state employment services. Civilian users include the American Petroleum Institute, Sears-Roebuck, and various insurance and utilities industries (Schmidt, 1988).

Validity generalization concepts have been incorporated in psychological texts (Anastasi, 1982), and in the Standards for Educational and Psychological Testing (APA, 1974). The Uniform Guidelines on Employee Selection Procedures issued in 1978 (Uniform Guidelines, 1978) are more accepting of validity generalization than the previous (1970) guidelines, and "spell out more precisely the methods necessary to "borrow validity" or generalize validity results to similar jobs" (Baker & Terpstra, 1982, p 603). Baker and Terpstra specify two court cases in which generalized validities have been accepted as evidence to refute racial and sex discrimination charges: Friend v. Leidinger, 1977, and Pegues v. Mississippi State Employment Service, 1980. It is thus not unreasonable to expect that future revisions of the guidelines will include an even broader acceptance of validity generalization procedures.

Statement of Purpose

This study investigated characteristics of ASVAB validity data via Schmidt and Hunter's meta-analytic procedures. In light of the research reviewed, it follows that General Cognitive Ability moderates the relationship between job knowledge and

job performance. Also, verbal and numerical skills constitute the major components of General Cognitive Ability. The AFQT, containing two verbal subtests (WK and PC), plus two numerical subtests (AR and NO), may be considered to represent a general cognitive measure.

Hunter's work in particular (1983a, 1983b, 1984, 1985) has demonstrated higher predictive validities for General Cognitive Ability than for the regression-based ASVAB selector composites. This study extends Hunter's work by attempting to demonstrate that the AFQT, as a general cognitive measure intrinsic to the ASVAB, is a valid predictor of Air Force technical training course success. In line with Hunter's results, the general cognitive measure (AFQT) was expected to improve prediction beyond that given by the four selector AIs (Mechanical, Administrative, General, and Electronic). Furthermore, it was expected that the improvement in prediction would be greater for women than for men. Females have been shown to score somewhat higher than men on cognitive ability measures, thus the females' validities were expected to raise when the AFQT alone was used as the predictor. The AFQT is not based on specialized experience to the extent of the other composites. For example, general cognitive measures (as with the AFQT) lack mechanically-related items, on which men tend to score higher than women.

Hypotheses

The formal hypotheses to be tested are:

1. Ho: $f_{AFQT} = 0$

   H_1: $f_{AFQT} > 0$ for all occupations.

The validity of the AFQT for predicting Air Force technical training success is essentially zero, for all job/training-course combinations.

2. Ho: $f_{AFQT} = f_M$

   H_1: $f_{AFQT} > f_M$ for Mechanical occupations.

AFQT validity will be essentially equal to (will not exceed) validity of the M composite for predicting training success in mechanical occupations.

3. Ho: $f_{AFQT} = f_A$

   H_1: $f_{AFQT} > f_A$ for Administrative occupations.

AFQT validity will be essentially equal to (will not exceed) validity of the A composite for predicting training success in administrative occupations.

4. Ho: $\rho_{AFQT} = \rho_{G}$

   H_1: $\rho_{AFQT} > \rho_{G}$ for General occupations.

AFQT validity will be essentially equal to (will not exceed) validity of the G composite for predicting training success in general occupations.

5. Ho: $\rho_{AFQT} = \rho_{E}$

   H_1: $\rho_{AFQT} > \rho_{E}$ for Electronic occupations.

AFQT validity will be essentially equal to (will not exceed) validity of the E composite for predicting training success in electronic occupations.

6. Ho: $\rho_{AFQT.F} - \rho_{AI.F} = \rho_{AFQT.M} - \rho_{AI.M}$

   H_1: $\rho_{AFQT.F} - \rho_{AI.F} > \rho_{AFQT.M} - \rho_{AI.M}$

Any incremental validity observed due to the AFQT will be equal for females and males, for all occupations.

CHAPTER II

METHODS

The Instrument

Current ASVAB versions (forms 8 through 14) consist of 334 items in 10 subtests. The ASVAB is group administered by trained DoD personnel in Military Entrance Processing Stations and their satellites (about 1,000 locations), and is also administered at approximately 14,000 high schools nation wide. The volume of tests given is approximately 2.3 million per year.

ASVAB subject areas were chosen according to their validity for predicting training success in each of the military services. The content of subtests of forms 8 through 14 is shown in Table 4. Total testing time is 144 minutes.

Subjects

Data were obtained from an AFHRL validation study performed by Wilbourn, Valentine, and Ree (1984). The subjects for their research consisted of 29,619 males and females, aged 17 to 24. A description of these subjects is given in Table 5. The subjects were all first-term Air Force enlistees, most of whom were high school graduates. They were tested on ASVAB forms 8, 9, and 10 between October, 1980, and March, 1982.

The subject group was restricted in range at the low score

Table 4

ASVAB Subtest Content

Subtest	Content	Number of Questions
General Science (GS)	general science, including biology and physics	25
Arithmetic Reasoning [a] (AR)	arithmetic word problems	30
Word Knowledge [a] (WK)	selecting synonyms	35
Paragraph Comprehension [a] (PC)	ability to understand a written text	15
Numerical Operations [a] [b] (NO)	simple, arithmetic calculations	50
Coding Speed [b] (CS)	substituting numeric codes for verbal material	84

Subtest	Content	Number of Questions
Auto and Shop Information (AS)	automobile and tool-usage knowledge	25
Mathematics Knowledge (MK)	calculations including algebra, geometry, and elementary trigonometry	25
Electronics Information (EI)	electrical principles and terminology	20

[a] AFQT subtests.

[b] Speeded subtests.

Table 5

Description of Subjects by Gender and Ethnicity

	N	Percent
White males	21,544	72.8
White females	2,702	9.1
Black males	4,040	13.6
Black females	590	2.0
Other races	733	2.5
Totals	29,619	100.0
Total white	24,256	81.9
Total black	4,630	15.6
Other races	733	2.5
Totals	29,619	100.0
Total males	26,259	88.7
Total females	3,360	11.3
Totals	29,619	100.0

levels to the extent that only those passing certain cutoff scores gained enlistment. (Selection criteria included the AFQT score, the sum of the four AIs, and the General AI score). An additional restriction factor curtailed the technical school samples at the upper score levels. Because each technical school maintains a cutoff score on an appropriate selector AI, higher-scoring subjects were assigned to schools with higher eligibility requirements.

The Criterion

The validation criterion used in the study was technical training course final grades. The courses included were the 70 courses which had at least 100 graduates. Data were computed on the subgroups which had at least 25 individuals. Technical school grades generally range between 70 (passing) and 100. Approximately four percent of enlistees do not pass the training courses. Attrition thus further restricted the sample range. As previously discussed, training school success is a highly appropriate criterion, as it directly predicts job performance (Hunter, 1984).

Procedures

Meta-analytic procedures were used to combine validities given in the Wilbourn, et al. data, across the four AIs and the

AFQT, for the entire group, and separately for males and females.

Meta-analytic procedures of Schmidt and Hunter utilize corrected correlations, however, no procedure is presently available for estimating the standard errors of corrected correlations. For this reason, the original data were combined twice; first, as uncorrected raw data; secondly, as corrected for range restriction and unreliability. Confidence intervals were then built around the uncorrected validities, and corrected data were applied to adjust the endpoints of the intervals (Hunter, Schmidt, & Jackson, 1982). The corrected confidence intervals were compared to determine overall effects. Detailed procedural steps were as follows:

For the corrected group:

1. correct validity coefficients for range restriction;

2. correct values obtained in step 1 for unreliability;

For both corrected and uncorrected groups:

3. transform validities to Fisher's z values;

4. weight and combine validities (z's) for total, male, and female groups, within each job category (M, A, G, and E);

5. estimate mean and standard deviation for each group from step 4;

6. convert z values to r's;

For computing confidence intervals:

7. compute confidence intervals around uncorrected

validities, correcting the endpoints for range

restriction and unreliability;

8. compare confidence intervals to determine whether they

overlap.

Reliability measures were obtained from a recent AFHRL

study (Palmer, Hartke, Ree, Welsh, & Valentine, 1988). The

correction calculations were obtained from Hunter, et al. (1982),

and are given in Appendix A. Fisher's z transformation was used

to convert the data into a linear scale. z values were obtained

from a transformation table (Agresti & Finlay, 1986).

The z transformation was used in this study despite the fact

that Hunter and Schmidt prefer not to utilize z's. Hunter, et

al. (1982) state that the z transformation produces positive bias

in obtained correlations. As evidence, they cite a validity

generalization study for computer programmers (Schmidt, Gast-

Rosenberg, & Hunter, 1980), in which average validities were

larger by about .07 due to the z transformation. They add,

however,

> The fact that the Fisher z transformation leads to bias in
>
> averaging correlations does not mean that there is any
>
> problem with the transformations in the context for which it
>
> was invented, i.e., confidence intervals. The Fisher z
>
> transformation expands large correlations relative to small
>
> ones, which causes the confidence interval around large

> correlations to be smaller than those around small
>
> correlations. However, the confidence intervals around
>
> large correlations should be smaller because sampling error
>
> is smaller. Thus the expansion has the desired effect on
>
> confidence intervals. (p 42)

Final results were returned to r values for interpretation.

Weighting procedures were used in step 4 for combining the validities, and in step 7 for computation of confidence intervals because the stability of a correlation is solely dependent upon the sample size. Weighting validities by the Ns of the samples gave more weight to correlations based on larger samples. Weighting the correlations was particularly important in order to allow gender comparisons. The Ns of female samples are typically smaller than those of male samples (see Appendix B for sample data).

Correcting for range restriction required three values: 1. standard deviations of the predictor variable (ASVAB scores), which represents the restricted sample; 2. the correlations of the predictor with the criterion (final school grades); and 3. the standard deviation of the unrestricted sample. This value (28.29) represents the standard deviation of any complete percentile scale (Netter, Wasserman, & Whitmore, 1988).

Reliabilities corrected included only those of the predictor. As is often the case, the criterion reliabilities

are difficult to estimate (Hunter, et al., 1982). The corrected values are, as a result, more conservative than if both predictor and criterion reliabilities were corrected. Additional uncorrectable factors, such as variations in technical school lengths, grading systems, and numbers of students, lend conservatism to the corrected values.

Analyses of the corrected data were performed via computer analysis, utilizing SPSS[X] procedures (SPSS Inc., 1986).

For the final step, evaluating the data for overall effects, 95 percent confidence intervals were formed around obtained values. Based on the methods of Hunter, et al. (1982), confidence intervals avoid problems of accumulating Type I and Type II errors. The endpoints of the intervals were corrected according to the formulae described in Appendix A. Overlapping confidence intervals were taken as indications that the values did not differ. The reader should be aware that the procedure of correcting the endpoints of confidence intervals can create nonsymmetrical intervals.

CHAPTER III

RESULTS

Tabled data are presented to four decimal places, for two
reasons. First, all calculations were made at this level of
precision. Secondly, rounding to two decimal places, as most
validity data are present ﹅ in the literature, obscured small
differences which were consi..ered important.

Table 6 presents the validities (mean r) as combined and
corrected through meta-analysis. Also shown are the 95-percent
confidence intervals computed around each validity coefficient.
Comparisons were made of validities resulting from prediction
by composites, versus validities resulting from prediction by the
AFQT for total, male, and female groups within each composite
category (M, A, G, and E). Comparisons which identified
significantly ($p \leq .05$) different validities (95 percent
confidence intervals did not overlap) are identified, with the
direction of the difference indicated by "<" or ">".

In Table 7, the proportion of variance accounted for by each
mean corrected validity coefficient is presented. The values in
the difference column ($r_c^2 - r_a^2$) indicate the amount of change in
prediction as a result of utilizing each alternative (composite
or AFQT). Alternatives with significantly higher validities are
indicated by the "<" or ">" comparison.

Table 6

Corrected Validities and Confidence Intervals for Comparison

Conditions

		Composite				AFQT	
		Mean r	95% CI			Mean r	95% CI
M	Total	.8330	.8189 to .8477	=		.8115	.7981 to .8287
	Males	.8475	.8314 to .8616	=		.8349	.8214 to .8503
	Females	.8305	.6509 to .9105	=		.8103	.7134 to .8702
A	Total	.4325	.3701 to .4859	<		.8561	.8331 to .8768
	Males	.5035	.4359 to .5632	<		.8689	.8392 to .8882
	Females	.2552	.1046 to .3651	<		.8655	.8067 to .9018
G	Total	.9142	.9030 to .9202	>		.8515	.8416 to .8660
	Males	.9220	.9035 to .9227	>		.8561	.8409 to .8675
	Females	.9510	.9124 to .9643	>		.8896	.8282 to .9044
E	Total	.9239	.9002 to .9218	>		.8384	.8279 to .8673
	Males	.9268	.9241 to .9430	>		.8589	.8422 to .8805
	Females	.9674	.7872 to .9887	=		.7855	.6451 to .9113

Table 7

Estimates of Proportion of Variance Accounted for by the

Composites versus the AFQT

		r^2 Composite		r^2 AFQT	$r^2_c - r^2_a$ [b]
M	Total	.6939	=[a]	.6585	.0354[a]
	Males	.7183	=[a]	.6970	.0212[a]
	Females	.6897	=[a]	.6566	.0331[a]
A	Total	.1870	<	.7329	-.5409
	Males	.2535	<	.7550	-.5015
	Females	.0651	<	.7491	-.6940
G	Total	.8358	>	.7250	.1108
	Males	.8501	>	.7329	.1172
	Females	.9044	>	.7914	.1130
K	Total	.8536	>	.7029	.1327
	Males	.8589	>	.7377	.1212
	Females	.9359	=[a]	.6170	.3189[a]

[a] Confidence intervals indicate no significant difference between the composite versus the AFQT for these groups, thus the r^2 differences may be interpreted as zero.

[b] Difference between r^2 composite and r^2 AFQT.

Null hypothesis number one was rejected. The alternate hypothesis was supported, which proposed that the AFQT would demonstrate validity for predicting training success for all occupations. AFQT validity was supported for all job groups, and independently for male and female subgroups. AFQT validities range from approximately .79 to .89, thus the AFQT accounts for 61 to 79 percent of the criterion variance.

Null hypothesis number two was not rejected. The alternate hypothesis proposed that the AFQT validity would be significantly higher than the validity of the Mechanical (M) composite for predicting training success for M jobs, but this was not supported. Analyses of the confidence intervals indicate equal validities for the M composite and the AFQT.

Null hypothesis three was rejected. Alternate hypothesis three, that the AFQT would have higher validity than the Administrative (A) composite, is clearly supported by the data. The AFQT accounts for about 73 to 75 percent of the variance (about 50 to 69 percent more than the A composite). Inspection of Table 6 shows substantial increases in validity due to the AFQT, but particularly for females.

According to null hypotheses four and five, the General (G) and Electronic (E) composites should have validities equal to the AFQT validity for predicting training success for G and E jobs respectively. The null hypotheses were not rejected in these

cases. Higher validities are indicated for the G and E composites than for the AFQT in all cases but one. That is, the data indicate composite and AFQT validities are equal for predicting training success only for E jobs, and only for the female group. For all other G and E conditions, the composites significantly outperform the AFQT, accounting for about 11 to 13 percent more variance than the AFQT (see Table 7).

According to alternate hypothesis six, any incremental validity seen due to the AFQT rather than the specific composites would be greater for females than males. Only in the A job categories were validities higher for the AFQT, and in these cases, females did show a significantly larger increase than males. For females, the increase (the difference between AFQT and A r-square values) was .6940. For males, the increase was .5015. A comparison of the confidence intervals (see Table 6) shows that the composite confidence interval for females lies below that for males, with no overlap. The conclusion drawn is that the increase in validity is clearly greater for females than for males.

CHAPTER IV

DISCUSSION

For predicting success in the mechanical jobs, M and AFQT validities are not significantly different, statistically speaking. However, had the mean validities for M been significantly higher, the incremental validity provided by M would range from .0212 to .0354 (r^2 values). These values would provide seemingly small advantages. With large samples, however, even small differences can be important. Small amounts of predictive validity can be advantageous when applied toward the classification of a large number of individuals. Practically speaking, then, these differences would support the continued use of the M composite for predicting success in M-type jobs. The AFQT (or any other composite) should not be substituted for M unless clearly shown to be a superior predictor over the M composite. Substantial differential validity favors the G and E composites for predicting success in general and electronic occupations, and favors the AFQT for predicting success in administrative occupations.

It is appropriate at this point to compare these results to those found by Hunter. In Hunter's (1983b) analysis, he included a technical composite, consisting of EI plus MC, as part of his measure of General Cognitive Ability (see figure 1). He stated

that "a measure of Technical Aptitude . . . . augments the

measure of General Cognitive Ability" (Hunter, et al., 1985,

p viii). In the present study, the AFQT has been cast in the

role of a measure of General Cognitive Ability. However, the

AFQT contains no Technical Aptitude subtests. In addition,

Hunter's General Cognitive Ability composite excluded speeded

subtests, but the AFQT contains a half-weighted value of NO.

Speeded tests such as NO commonly have low validity (Hunter,

et al., 1985; Thorndike, 1986). These differences between the

measures of General Cognitive Ability used in the two studies

could help to explain why Hunter's analysis showed less

differential validity than is observed in the present study.

Another difference between the two studies is the number of

subtests contained in the General Cognitive Ability measure.

Hunter included six reliable subtests, but this study used only

three subtests with equally acceptable reliability (reliability

of NO is low in comparison with other subtests). Because

reliability of measurement increases as the number of reliable

components increases, error of measurement may be less for

Hunter's study.

The AFQT was chosen to represent General Cognitive Ability

for this meta-analysis for two reasons. First, its verbal and

mathematic subtests (respectively, WK and PC; AR and NO) are

representative of the major areas on which most batteries

measuring General Cognitive Ability concentrate. (Many educational batteries, for example, rely on only verbal and quantitative aptitude measures [Hunter, et al., 1985]). Secondly, because AFQT scores are currently computed for ASVAB examinees, transition from using a less valid composite to the AFQT could be accomplished with relative ease.

Recent research (Jones, 1988) indicates that the AFQT may not, in fact, be the best measure of g, or General Cognitive Ability obtainable via the ASVAB subtests. Jones determined the g loadings of the 10 ASVAB subtests through a principle components analysis, then ranked the subtests' g loadings, and determined the Spearman rank-difference correlation coefficient. Her data base for the principle components analysis was the 1980 Profile of American Youth (N = approximately 12,000). Data for the correlational computation were chosen from an Air Force accessions validity study (Wilbourn, Valentine, & Ree, 1984). (Total N for Jones' study = 24,482; N of her samples ranged from 107 to 4,673). She found that the subtests' g loadings are positively correlated with their average corrected validity coefficients (Spearman's rho = .75). Jones' validity data, and the data used in this meta-analysis are from the same source. As would be expected, the results are rather consistent, regardless of the analysis performed.

Table 8 presents the subtests and their g loading ranks as

determined by Jones, as well as the subtest configurations of the composites. A number of points may be made relative to these results.

First, it becomes clear that a better measure of g than the AFQT configuration is available. The AFQT subtests WK and AR appear highly g loaded, as would be expected. However, PC and NO could be replaced with GS and MK, to utilize the four most highly ranked g-loaded subtests. Secondly, both speeded subtests of the ASVAB (NO and CS) rank comparatively low in terms of g (eighth and tenth). That these subtests have quite low g ranks, as well as low validities merits their exclusion from a composite intended to measure General Cognitive Ability.

Jones' results are helpful in explaining the results of the present study. Most striking is the comparison of the g-loading ranks of the E composite versus the AFQT. The median g-loading rank of the composite is approximately 3, compared to a median rank of the AFQT of about 5. All E-composite subtests, including the specifically electronic subtest EI, rank quite high in terms of g, and in fact, appear to measure g better than the AFQT. The basic premise of the present study, that highly g loaded composites (or, good measures of General Cognitive Ability) should have higher predictive validities than specialized composites, is supported by Jones' results (see Table 8). If the E composite were, in fact, a better measure of g than the AFQT,

Table 8

g-loading Rankings of ASVAB Subtests

g-loading Rank	g Loadings	Subtest	M	A	G	E	AFQT
1	.8831	GS	X			X	
2	.8792	WK		X	X		X
3	.8769	AR			X	X	X
4	.8291	MK				X	
5	.8220	EI				X	
6	.8170	PC		X	X		X
7	.7965	MC	X				
8	.7219	NO		X			X[b]
9	.6956	AS	X[a]				
10	.6367	CS		X			

[a] Double-weighted.

[b] Half-weighted.

the results obtained through the present meta-analysis would be expected, because the meta-analysis shows that E demonstrates higher validities than the AFQT.

This conclusion leads to a consideration of the G composite as compared to the AFQT, in terms of g loadings. As can be observed in Table 8, the G composite is actually a subset of the AFQT, differing only as to its exclusion of NO. Adding components to a measuring instrument will improve measurement only if the new components are reliable and valid. Addition of NO to the G-composite subtests, resulting in the AFQT configuration, would not be expected to improve measurement because NO is insufficiently reliable compared to the other subtests. If inclusion of NO decreases overall validity of the AFQT, we would expect the present results to reflect lower validities for the AFQT than for the G composite. This is, in fact, what is observed. AFQT validities range from .8515 to .8896 (mean r values) for the general occupations. According to the analyses, these values are significantly lower than G-composite validities (mean r ranges from .9142 to .9510). The G composite accounts for approximately 11 percent more variance than the AFQT (see Table 7).

A final observation relates to the A composite. The g-loading rankings and meta-analytic results again appear to coincide. Meta-analysis shows the AFQT to be a more valid

predictor than the A composite. The median g-loading rank of the composite is about 6, compared to about 5 for the AFQT. In comparison, both composites include WK and PC. The A composite, however, contains two speeded subtests, both of comparatively low g loading and validity. Three of the four A composite subtests rank among the lower half, in terms of g. Additionally, the AFQT contains AR, a fairly highly g-loaded subtest. These differences indicate that the AFQT is more highly g loaded than A, and therefore could be expected, as Jones' results indicate, to predict better than the A composite.

Speeded subtests have been considered problematic, particularly due to their characteristics of low validity (Thorndike, 1986) and low reliability (Wegner & Ree, 1986). Wegner and Ree (1986) investigated the effects of the speeded subtest, NO, upon the validity of the AFQT. Specifically, they attempted to identify ASVAB subtest composites alternative to the present AFQT that would contain no speeded subtests, and that would produce validities at least as high as the current AFQT. Utilizing ASVAB scores of 154,788 Air Force recruits, with technical training course grades as the validation criterion, they assessed the validities of 15 AFQT alternatives. All of the alternatives produced uncorrected validities higher than that of the current AFQT. The researchers proposed two possible reasons for this finding. First, the results could be due to effects of

the speeded subtest in the current AFQT. Secondly, validity coefficients may have been reduced due to curtailment of the sample as a result of initial selection on the basis of the General composite. It should be noted again that the G composite is a subset of the AFQT, and also, when analyzed as one of the 15 AFQT alternatives, the G composite produced the second-lowest validity coefficient.

Wegner and Ree identified three of the alternative AFQTs as being acceptable alternatives to the current configuration. One of these has been chosen to replace the current AFQT as of January 1, 1989. The replacement AFQT is composed of AR + MK + 2VE, where VE = WK + PC.

Table 9 shows the current and replacement AFQT compositions in relation to the g-loading ranks of their subtests. The replacement AFQT demonstrates higher validity (Wegner & Ree, 1986) and in relation to Jones' results, higher g loading than the current AFQT. In addition, the replacement produces minimal change in the enlistment qualifications of specific subgroups (females, blacks, and minorities).

A discrepancy is noted between the present meta-analytic results and those of Hunter, et al. (1985). Their results, based on numerous data sets, indicated that a Speed factor may contribute something to validity for clerical occupations above what is contributed by General Cognitive Ability. Their analysis

Table 9

g-loading Rankings of Current and Replacement AFQTs

g-loading Rank	g Loadings	Subtest	Current AFQT	Replacement AFQT
1	.8831	GS		
2	.8792	WK	X	X[a]
3	.8769	AR	X	X
4	.8291	MK		X
5	.8220	EI		
6	.8170	PC	X	X[a]
7	.7965	MC		
8	.7219	NO	X[b]	
9	.6956	AS		
10	.6367	CS		

[a] Double-weighted.

[b] Half-weighted.

utilizing occupational composites reported in the DoD Student
Testing Program (often called the High School ASVAB) resulted in
a Business and Clerical composite validity of .59 for
corresponding occupations, and only about .54 for other
occupations. (The Business and Clerical composite was composed
of WK + PC + CS + MK). Validity of General Cognitive Ability was
only .56 for business and clerical occupations. This meta-
analysis indicated opposite results. That is, the general
cognitive measure (AFQT) had higher validity than the clerical
composite (A) for administrative-type jobs.

Additionally, a meta-analysis of eight military studies
produced validity coefficients of a Speed factor for clerical
work of .43, versus .37 for other types of work (Hunter, et al.,
1985). (Note that these findings do not contradict Hunter's
previously described work based on the data of Sims and Hiatt
[1981], Maier and Grafton [1981], and Thorndike 7], on which
he based his path model. The path model did not utilize speeded
subtests [see figures 1 and 2]). The present meta-analytic
results suggest that speeded subtests decrease validity in
general, but that they may be of even greater detriment to the
business/clerical (administrative) occupations than others.

The reason for this discrepancy regarding the effects of
speeded subtests for predicting clerical work is not clear. A
possible explanation lies in the differences among the predictive

composites and occupational criteria used in the various studies.
The composite configurations used by Hunter and his associates
were occupational composites constructed to be relative to
civilian occupations (in reference to the high school composites)
and were composites used by various branches of the service. The
present study, however, analyzed only Air Force composites and
occupations. Further study is necessary to conclusively verify
this assumption.

Alternate hypothesis six proposed that the AFQT would
provide greater incremental validity for females than for males.
This was verified only for administrative occupations. Also, the
administrative occupational group was the only group for which
the AFQT validity was significantly higher than the composite
validity. If a different general cognitive measure were to be
employed, with higher g loading than the AFQT, we might expect
to see higher validities for the general cognitive measure as
well as higher increments for females, in all occupational
groups. Continued research is necessary, however, to investigate
possible explanations for the gender differences. Such research
might include investigation of the effects of the A composite's
speeded subtests.

Two possible explanations are proposed for the differential
validity observed in this meta-analysis: 1. The differences
accurately represent predictive abilities of the various

composites; 2. As a measure of General Cognitive Ability, the AFQT is insufficient.

In certain conditions, evidence of differential validity may be expected. For example, Hunter et al. (1985) stated:

. . . . for certain highly geometric professions in engineering and the graphic arts, Spatial Aptitude makes a contribution to job performance beyond that of General Cognitive Ability alone. Thus, there are some jobs in which people do build on specific learning which is measured by the specific aptitudes. (p 92)

They specify however, that this is not the case with Technical aptitude; the importance of a Technical aptitude factor is in its increment to the measurement of General Cognitive Ability. The aptitude areas measured by the ASVAB (verbal, quantitative, and technical) are relatively general, existing at the level of second-order factors, or group factors, according to Jensen (1986). Spatial aptitude has not proven useful as a group factor or as a component in group factors, for prediction of military training course success. In fact, a Spatial Perception subtest (SP), previously contained in the ASVAB, was dropped with ASVAB forms 8, 9, and 10. Hunter (1983b) indicated that, in the presence of good Technical and Quantitative measures, SP did not raise validity meaningfully.

The conclusion for this meta-analysis relies on the work of

Jones (1988) to support the second explanation of the differential validity observed. As major points: 1. General Cognitive Ability as represented by the AFQT is more valid as a predictor than the less g-loaded Administrative composite; 2. The Electronic composite represents g more strongly than the AFQT, thus resulting in higher validities for E; 3. The AFQT demonstrates lower validities than the General composite, which is presumably due to the AFQT's inclusion of the speeded, low g-ranked subtest, NO; 4. g loadings of the AFQT versus the Mechanic composite are not clearly determinable by visual inspection. However, results indicate little differential validity (.0212 to.0354) and these differences are not statistically significant.

Recommendations

Predictive validity of the ASVAB will be improved by replacing the A composite with a more valid measure. The current AFQT is more valid than A, thus it would be a useful replacement, and would provide an easy transition because it is currently scored. The G composite may also prove to be a useful replacement for A. While the AFQT appears to work equally well for all job categories, the G composite is more valid than the AFQT for G-type jobs (see Table 6), thus it may be more useful than the AFQT for A-type jobs as well. Other subtest

configurations should be investigated, however, if the goal is to
provide the best predictive measure for administrative
occupations. Neither the AFQT nor the G composite should be
assumed the best possible configuration available for replacing
the A composite.

The speeded subtests, NO and CS, have low validities,
reliabilities, and g loadings. Problems presumed due to its
speeded component have resulted in reconfiguration of the AFQT.
Further investigations of the effects of these subtests may
indicate merit in dropping them entirely. As of January 1,
1989, the only Air Force composite to utilize NO and/or CS will
be the A composite. Research could be performed to determine
whether validity of the A composite may be improved simply by
dropping the problematic speeded subtests.

A meta-analysis similar to the present study should be
performed, based on data from the replacement AFQT, when it
becomes available after January 1, 1989. Additionally, similar
meta-analyses would be highly informative in four variations:

1. Employ a General Cognitive Ability composite that is
   more highly g loaded than the current AFQT, possibly
   utilizing the four subtests with highest g-loading ranks
   (see Table 8).

2. Employ a General Cognitive Ability composite which
   includes measures of Technical aptitude, such as the EI

and MC subtests.

3. Evaluate effects of gender and ethnicity when EI and MC
   are added to the measure of General Cognitive Ability.

4. Evaluate validity of the G composite as a General
   Cognitive Ability measure, including a comparison of the
   effects of G versus the current AFQT, to assess the
   effects of the speeded AFQT component, NO.

REFERENCES

Agresti, A., & Finlay, B. (1986). Statistical Methods for the Social Sciences. San Francisco: Dellen Publishing Company.

Anastasi, A. (1982). Psychological Testing. (5th Ed.) New York: Macmillan.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education (1974). Standards for Educational and Psychological Tests. Washington, D. C.: American Psychological Association.

Baker, D. D., & Terpstra, D. E. (1982). Employee selection: must every job be validated? Personnel Journal, 61, 602-605.

Bartlett, C. J., Bobko, P., Mosier, S. B., & Hannon, R. (1978). Testing for fairness with a moderated multiple regression strategy: an alternative to differential analysis. Personnel Psychology, 31, 233-241.

Block, J. B. (1968). Hereditary components in the performance of twins on the WAIS. In S. G. Vandenberg (Ed.), Progress in Human Behavior Genetics. Baltimore: The Johns Hopkins University Press.

Bock, R. D., & Mislevy, R. J. (1981). Data Quality Analysis of the Armed Services Vocational Aptitude Battery. Chicago, IL: National Opinion Research Center.

Coggin, T. D., & Hunter, J. E. (1983). Problems in measuring the

quality of investment information: the perils of the
information coefficient. Financial Analysts Journal,
May/June, 1-10.

DuBois, P. H. (1970). A History of Psychological Testing.
Boston: Allyn and Bacon.

Eysenk, J. J., & Barrett, P. (1985). Psychophysiology and the
measurement of intelligence. In C. R. Reynolds & V. Wilson
(Eds.), Methodological and Statistical Advances in the Study
of Individual Differences. New York: Plenum.

Fisher, C. D., & Gittelson, R. (1983). A meta-analysis of the
correlates of role conflict and ambiguity. Journal of
Applied Psychology, 68, 320-333.

Friedman, D., Streicher, A., Wing, H., Grafton, F., & Mitchell,
K. (1983). Reliability of scores for fiscal year 1981 Army
applicants: Armed Services Vocational Aptitude Battery forms
8, 9, and 10. Alexandria, VA: U.S. Army Research Institute
for the Behavioral and Social Sciences.

Ghiselli, E. E. (1966). The Validity of Occupational Aptitude
Tests. New York: John Wiley & Sons.

Ghiselli, E. E. (1973). The validity of aptitude tests in
personnel selection. Personnel Psychology, 26, 461-477.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of
research. Educational Researcher, 5, 3-8.

Glass, G. V. (1977). Integrating findings: the meta-analysis of

research. Review of Research in Education, 5, 351-379.

Gottfredson, L. S. (1986). Societal Consequences of the g factor
in employment. Journal of Vocational Behavior, 29, 379-410.

Graham, J. R., & Lilly, R. S. (1984). Psychological Testing.
Englewood Cliffs, NJ: Prentice-Hall, Inc.

Haier, R. J., Robinson, D. L., Braden, W., & Williams, D. (1983).
Electrical potentials of the cerebral cortex and psychometric
intelligence. Personality and Individual Differences, 4,
591-599.

Hawk, J. (1986). Real world implications of g. Journal of
Vocational Behavior, 29, 411-414.

Helme, W. E., Gibson, W. A., & Brogden, H. E. (1957). An
Empirical Test of Shrinkage Problems in Personnel
Classification Research. Personnel Board Technical Research
Note 84.

Hunter, J. E. (1983a). The dimensionality of the General
Aptitude Test Battery (GATB) and the dominance of the general
factor over specific factors in the predictions of job
performance for USES. (Test Research Report No. 44)
Washington, D. C.: U.S. Department of Labor, U.S. Employment
Services.

Hunter, J. E. (1983b). Validity generalization of the ASVAB:
Higher validity for factor analytic composites. Rockville,
MD: Research Applications, Inc.

Hunter, J. E. (1984). The validity of the ASVAB as a predictor of civilian job performance. Rockville, MD: Research Applications, Inc.

Hunter, J. E. (1985). Validity generalization of the ASVAB: preliminary report. Rockville, MD: Research Applications, Inc.

Hunter, J. E., Crosson, J. J., & Friedman, D. H. (1985). The validity of the Armed Services Vocational Aptitude Battery (ASVAB) for civilian and military job performance. Rockville, MD: Research Applications, Inc.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. Psychological Bulletin, 96, 72-98.

Hunter, J. E., & Schmidt, F. L. (1982). Ability tests: Economic benefits versus the issue of fairness. Industrial Relations, 21, 293-308.

Hunter, J. E., & Schmidt, F. L. (1983). Quantifying the effects of psychological interventions on employee job performance and work-force productivity. American Psychologist, 38, 473-478.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: a comprehensive review and analysis. Psychological Bulletin, 86, 721-735.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). Meta-

Analysis: Cumulating Research Findings Across Studies.
Beverly Hills, CA: Sage Publications.

Jensen, A. R. (1980). Bias in Mental Testing. New York:
Macmillan Publishing Co., Inc.

Jensen, A. R. (1983). Effects of inbreeding on mental ability
factors. Personality and Individual Differences, 4, 71-87.

Jensen, A. R. (1986). g: artifact or reality? Journal of
Vocational Behavior, 29, 310-331.

Jones, G. E. (1988). Investigation of the efficacy of general
ability versus specific abilities as predictors of
occupational success. Unpublished master's thesis, St.
Mary's University, San Antonio, TX.

Kass, R. A., Mitchell, K., Grafton, F., & Wing, H. (1982).
Factor structure of the Armed Services Vocational Aptitude
Battery (ASVAB) forms 8, 9, and 10: Army applicant sample.
Alexandria, VA: U.S. Army Research Institute for the
Behavioral and Social Sciences.

Maccoby, E. E. (Ed.) (1966). The Development of Sex Differences.
Stanford, CA: Stanford University Press.

Maccoby, E. E., & Jacklin, C. N. (1974). The Psychology of Sex
Differences. Stanford, CA: Stanford University Press.

Maier, M. H., & Grafton, F. C. (1981). Aptitude composites for
ASVAB forms 8, 9, and 10. (Research Report 1308) Alexandria,
VA: U.S. Army Research Institute for the Behavioral and

Social Sciences.

McCarthy, D. (1954). Language development in children. In L. Carmichael (Ed.) Manual of Child Psychology. (2nd Ed.) New York: Wiley.

McNemar, Q. (1964). Lost: Our intelligence? Why? American Psychologist, 19, 871-882.

Nairn, A. & Associates (1980). The reign of ETS: The corporation that makes up minds. Washington, D. C.: Nader.

Nagoshi, C. T., & Johnson, R. C. (1986). The ubiquity of g. Personality and Individual Differences, 7, 201-207.

Netter, J., Wasserman, W., & Whitmore, G. A. (1988). Applied Statistics. (3rd Ed.) Boston: Allyn and Bacon.

Palmer, P., Hartke, D. D., Ree, M. J., Welsh, J. R., & Valentine, L. D. Jr. (1988). Armed Services Vocational Aptitude Battery (ASVAB): Alternate form reliability (forms 8, 9, 10, and 11). (Report No. AFHRL TP-87-48) Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. Journal of Applied Psychology, 65, 373-406.

Premack, S., & Wanous, J. P. (1985). Meta-analysis of realistic job preview experiments. Journal of Applied Psychology, 70, 706-719.

Ree, M. J., Mathews, J. J., Mullins, C. J., & Massey, R. H. (1982). Calibration of Armed Services Vocational Aptitude Battery forms 8, 9, and 10. (Report No. AFHRL TR-81-49) Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Reilly, R. R., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. Personnel Psychology, 35, 1-62.

Rosenthal, R. (1978). Combining results of independent studies. Psychological Bulletin, 85, 185-193.

Sackett, P. R., Schmitt, N., Tenopyr, M. L., Kehoe, J., & Zedeck, S. (1985). Commentary on forty questions about validity generalization and meta-analysis. Personnel Psychology, 38, 697-798.

Sims, W. H., & Hiatt, C. M. (1981). Validation of the Armed Services Vocational Aptitude Battery (ASVAB) forms 6 and 7 with application to ASVAB forms 8, 9, and 10. Alexandria, VA: Marine Corps Operations Analysis Group, CNS 1160, Center for Navy Analysis.

Schmidt, F. L. (1988). Validity generalization and the future of criterion-related validity. In H. Wainer & H. I. Braun (Eds.) Test Validity. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schmidt, F. L., Gast-Rosenberg, I., & Hunter, J. E. (1980). Validity generalization: results for computer programmers.

Journal of Applied Psychology, 58, 5-9.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 62, 529-540.

Schmidt, F. L., & Hunter, J. E. (1978). Moderator research and the law of small numbers. Personnel Psychology, 31, 215-232.

Schmidt, F. L., & Hunter, J. E. (1981). Employment testing. Old theories and new research findings. American Psychologist, 36, 1128-1137.

Schmidt, F. L., & Hunter, J. E. (1984). A within setting empirical test of the situational specificity hypothesis in personnel selection. Personnel Psychology, 37, 317-326.

Schmidt, F. L., Hunter, J. E., & Caplan, J. R. (1981). Validity generalization results for two job groups in the petroleum industry. Journal of Applied Psychology, 66, 261-273.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Task differences as moderators of aptitude test validity in selection: A red herring. Journal of Applied Psychology, 66, 166-185.

Schmidt, F. L., Hunter, J. E., & Urry, V. W. (1976). Statistical power in criterion-related validity studies. Journal of Applied Psychology, 61, 473-485.

Schmidt, F. L., Hunter, J. E., Outerbridge, A. N., & Trattner, M. H. (1986). The economic impact of job selection methods on

size, productivity, and payroll costs of the federal work force: an empirically based demonstration. Personnel Psychology, 39, 1-29.

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Hirsch, H. R. (1985). Forty questions about validity generalization and meta-analysis. Personnel Psychology, 38, 697-798.

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Shane, G. S. (1979). Further tests of the Schmidt-Hunter Bayesian validity generalization procedure. Personnel Psychology, 32, 257-281.

Schmidt, F. L., Pearlman, K., & Hunter, J. E. (1980). The validity and fairness of employment and educational tests for Hispanic Americans: a review and analysis. Personnel Psychology, 33, 705-724.

Smith, I. M. (1948). Measurement of spatial ability in school pupils. Occupational Psychology, 22, 150-159.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. American Journal of Psychology, 15, 210-293.

Spearman, C. (1927). The Abilities of Man. New York: Macmillan.

SPSS Incorporated. (1986). SPSS[x] User's Guide (2nd Ed.). Chicago, IL: Author.

Stacey, W. D., Weismuller, J. J., Barton, B. B., & Rogers, C. R. (1974). CODAP: Control Card Specifications for the Univac

1108. (AFHRL TR-74-84) Lackland Air Force Base, TX: Air

    Force Human Resources Laboratory.

Tambs, K., Sundet, J. M., & Magnus, P. (1984). Heritability

    analysis of the WAIS subtests. A study of twins.

    Intelligence, 8, 283-293.

Thorndike, R. L. (1957). The optimum test composites to predict

    a set of criteria. (Report No. AFPTRC-TN 57-103) Lackland Air

    Force Base, TX: Air Force Personnel Training and Research

    Center.

Thurstone, L. L. (1938). Primary Mental Abilities. Chicago:

    University of Chicago Press.

Thurstone, L. L., & Thurstone, T. G. (1941). Factorial Studies

    of Intelligence. Chicago: University of Chicago Press.

Tyler, L. E. (1986). Back to Spearman? Journal of Vocational

    Behavior, 29, 445-450.

Uniform Guidelines on Employee Selection Procedures (1978).

    Federal Register, 43, 38290-38309.

U.S. Department of Defense (1984). Armed Services Vocational

    Aptitude Battery (ASVAB): Test Manual. North Chicago, IL:

    U.S. Military Entrance Processing Command.

U.S. Naval Personnel Research Activity (1981). Basic Test

    Battery Validity Report for 89 Class "A" School Samples.

    San Diego, CA: Author.

Wegner, T. G., & Ree, M. J. (1986). Alternative Armed Forces

Qualification Test Composites. (Report No. AFHRL TP-86-27) Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Wilbourn, J. M., Valentine, L. D. Jr., & Ree, M. J. (1984). Relationships of the Armed Services Vocational Aptitude Battery (ASVAB) forms 8, 9, and 10 to Air Force technical school final grades. (Report No. AFHRL TP-84-8) Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

APPENDIX A

COMPUTATIONAL FORMULAE FOR

META-ANALYTIC PROCEDURES

Formulae are taken from Hunter, Schmidt, & Jackson (1982).
Computations were performed with sample estimates of the
population parameters.

1. Correction for restriction in range:

$$U = \frac{S}{s}$$

Where: S = standard deviation of
percentile scores for
the unrestricted group
(population)

and: s = standard deviation of
the predictor (ASVAB
scores) given by
Wilbourn et al. (1984)

$$\mathscr{P}_1 = \frac{U \, \mathscr{P}_2}{\sqrt{(U^2 - 1)\mathscr{P}_2^2 + 1}}$$

Where: $\mathscr{P}_1$ = reference population
correlation

and: $\mathscr{P}_2$ = study (sample)
correlation

2. Correction for unreliability of the predictor:

$$r_c = \frac{r_{xy}}{\sqrt{r_{xx}}}$$

Where: r_{xy} = validity coefficient

and: r_{xx} = reliability coefficient
of the predictor

3. Estimating mean and variance (performed with r's as represented in Fisher's z form):

Mean: $\quad \bar{r} = \dfrac{\sum (N_i \, r_i)}{\sum N_i}$

Corresponding
   variance: $\quad s_r^2 = \dfrac{\sum [\, N_i \, (r_i - \bar{r})^2 \,]}{\sum N_i}$

4. Confidence intervals:

Step 1: Compute confidence intervals around uncorrected correlations.

Step 2: Apply formulae given in #1 and #2 to correct endpoints of obtained intervals for range restriction and unreliability.

APPENDIX B

RAW VALIDITY DATA

Table B-1

Raw Validity Data

MECHANICAL OCCUPATIONS

AFSC	Composite or AFQT	Total		Males		Females	
		N	r	N	r	N	r
36130	C	127	.37	125	.39		
36130	A	127	.44	125	.44		
42331	C	361	.33	308	.31	53	.20
42331	A	361	.26	308	.29	53	.42
42333	C	431	.41	363	.44	68	.05
42333	A	431	.27	363	.31	68	.22
42632	C	1238	.46	1079	.46	159	.19
42632	A	1238	.42	1079	.45	159	.33
42633	C	165	.43	146	.46		
42633	A	165	.42	146	.49		
42733	C	158	.31	140	.34		
42733	A	158	.27	140	.22		
42735	C	550	.27	489	.28	61	.21
42735	A	550	.41	489	.42	61	.31
47231	C	134	.52	119	.57		
47231	A	134	.22	119	.27		

AFSC	Composite or AFQT	Total		Males		Females	
		N	r	N	r	N	r
47232	C	135	.47	123	.48		
47232	A	135	.34	123	.34		
55130	C	151	.36	149	.36		
55130	A	151	.34	149	.32		
55230	C	100	.29	91	.21		
55230	A	100	.19	91	.23		
55232	C	115	.47	98	.44		
55232	A	115	.42	98	.44		
56631	C	172	.41	155	.42		
56631	A	172	.43	155	.51		
11430	C	111	.25	105	.27		
11430	A	111	.42	105	.42		
42731	C	322	.16	293	.19	29	.35
42731	A	322	.41	293	.39	29	.43
43130	C	155	.46	150	.45		
43130	A	155	.37	150	.38		
43131	C	2179	.47	2107	.47	72	.35
43131	A	2179	.46	2107	.47	72	.38
43132	C	2216	.49	2124	.50	92	.43
43132	A	2216	.44	2124	.46	92	.46
44330	C	248	.47	242	.46		

AFSC	Composite or AFQT	Total N	Total r	Males N	Males r	Females N	Females r
44330	A	248	.39	242	.37		
46330	C	217	.39	203	.41		
46330	A	217	.64	203	.66		

ADMINISTRATIVE OCCUPATIONS

AFSC	Composite or AFQT	Total N	Total r	Males N	Males r	Females N	Females r
60230	C	107	.26	75	.27	32	.28
60230	A	107	46	75	.47	32	.43
60231	C	107	.29	78	.31	29	.19
60231	A	107	.54	78	.53	29	.58
70230	C	1841	.15	1280	.20	561	.02
70230	A	1841	.38	1280	.20	561	.31
60530	C	166	.41	111	.42	55	.38
60530	A	166	.51	111	.53	55	.47
20731	C	138	.33	96	.39	42	.18
20731	A	138	.64	96	.65	42	.55
29333	C	132	.16	96	.22	36	.01
29333	A	132	.30	96	.33	36	.21
73230	C	679	.31	485	.33	194	.25
73230	A	679	.51	485	.53	194	.45

GENERAL OCCUPATIONS

AFSC	Composite or AFQT	Total		Males		Females	
		N	r	N	r	N	r
57130	C	817	.44	793	.44		
57130	A	817	.42	793	.42		
62230	C	448	.38	307	.37	141	.40
62230	A	448	.35	307	.33	141	.41
81130	C	4673	.45	4673	.45		
81130	A	4673	.42	4673	.42		
81132	C	1855	.49	1617	.49		
81132	A	1855	.46	1617	.46		
27630	C	248	.47	197	.50	51	.34
27630	A	248	.44	197	.48	51	.27
29130	C	348	.32	228	.31	120	.34
29130	A	348	.30	228	.27	120	.37
51130	C	192	.43	155	.41	37	.38
51130	A	192	.42	155	.43	37	.31
90230	C	608	.57	446	.59	162	.51
90230	A	608	.52	446	.53	162	.47
90430	C	121	.37	88	.35	33	38
90430	A	121	.37	88	.37	33	.36

AFSC	Composite or AFQT	Total N	Total r	Males N	Males r	Females N	Females r
90630	C	240	.38	177	.37	63	.42
90630	A	240	.39	177	.40	63	.36
91530	C	105	.37	73	.43	32	.24
91530	A	105	.38	73	.39	32	.40
98130	C	222	.32	171	.32	51	.40
98130	A	222	.32	171	.32	51	.30
55330	C	130	.59	106	.56		
55330	A	130	.58	106	.55		
20230	C	135	.45	102	.40	33	.57
20230	C	135	.40	102	.33	33	.56
25130	C	195	.34	142	.33	53	.29
25130	A	195	.33	141	.33	53	.29
64531	C	538	.35	419	.35	119	.38
64531	A	538	.36	419	.36	119	.40

ELECTRONIC OCCUPATIONS

AFSC	Composite or AFQT	Total N	Total r	Males N	Males r	Females N	Females r
40431	C	142	.60	117	.62	25	.48
40431	A	142	.53	117	.56	25	.35
42330	C	561	.55	488	.53	73	.38

(table continues)

AFSC	Composite or AFQT	Total		Males		Females	
		N	r	N	r	N	r
42330	A	561	.47	488	.46	73	.35
30333	C	113	.38	107	.37		
30333	A	113	.27	107	.27		
44530	C	210	.52	205	.50		
44530	A	210	.45	205	.43		
30430	C	219	.55	203	.56		
30430	A	219	.43	203	.46		
30434	C	366	.49	332	.48	34	.34
30434	A	366	.31	332	.34	34	.08
30534	C	237	.45	215	.43		
30534	A	237	.29	215	.28		
30630	C	144	.59	125	.57		
30630	A	144	.43	125	.44		
30632	C	138	.43	121	.45		
30632	A	138	.28	121	.33		
30730	C	180	.37	146	.40	34	.14
30730	A	180	.31	146	.34	34	.36
31630	C	328	.45	309	.47		
31630	A	328	.24	309	.29		
32130	C	127	.41	117	.45		

(table continues)

AFSC	Composite or AFQT	Total		Males		Females	
		N	r	N	r	N	r
32130	A	127	.29	117	.32		
32123	C	288	.49	262	.48	26	.54
32123	A	288	.43	262	.45	26	.38
32232	C	244	.49	222	.50		
32232	A	244	.36	222	.36		
32430	C	100	.46	93	.47		
32430	A	100	.35	93	.36		
32530	C	245	.41	236	.43		
32530	A	245	.26	236	.27		
32531	A	342	.40	314	.40	28	.33
32531	A	342	.34	314	.35	28	.35
32633	C	105	.54	101	.54		
32633	A	105	.54	101	.56		
32634	C	188	.36	167	.39		
32634	A	188	.29	167	.27		
32636	C	283	.48	254	.53	29	.03
32636	A	283	.31	254	.32	29	
32637	C	206	.41	190	.37		
32637	A	206	.19	190	.21		
32638	C	290	.54	271	.54		

(table continues)

AFSC	Composite or AFQT	Total		Males		Females	
		N	r	N	r	N	r
32638	A	290	.39	271	.39		
32830	C	351	.56	317	.55	34	.50
32830	A	351	.36	317	.39	34	.40
32831	C	297	.45	271	.47	26	.07
32831	A	297	.33	271	.38	26	.00
32833	C	244	.53	234	.55		
32833	A	244	.45	234	.48		
32834	C	218	.44	206	.44		
32834	A	218	.27	206	.29		

VITA

CENSUS: Nicole S. Stermer was born on September
                     14, 1958, in Fargo, North Dakota. Her
                     parents are Tony and Irene Schafer, of
                     Dickinson, North Dakota. She is married
                     with no children.

TRAINING: Nicole S. Stermer graduated from Dickinson
                     High School, Dickinson, North Dakota, in
                     May, 1976. She received her bachelor of
                     arts degree from the University of North
                     Dakota, Grand Forks, North Dakota, in
                     December, 1982. She has performed
                     graduate work in 1984 at Georgia College
                     in Milledgeville, Georgia, and in 1986
                     and 1987 at the University of Texas at
                     San Antonio, Texas.

EXPERIENCE: Publications and other achievements:

                     In 1982, she completed an undergraduate
                     thesis in psychology. From 1983 until
                     1985, she taught mentally retarded adults
                     for the Houston County Association for
                     Exceptional Children, in Warner Robins,
                     Georgia. In 1985, she was commissioned
                     in the U.S. Air Force, and currently
                     serves as a behavioral scientist.

PERMANENT ADDRESS: 612 Meadow Gate
                     Converse, Texas

TYPIST: Nicole S. Stermer